IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 1, 2020, accepted February 20, 2020, date of publication February 28, 2020, date of current version March 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2977117

Data Anonymization Using Pseudonym System to Preserve Data Privacy

SHUKOR ABD RAZAK, NUR HAFIZAH MOHD NAZARI, AND ARAFAT AL-DHAQM
School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Skudai 81310, Malaysia
Corresponding author: Shukor Abd Razak (shukorar@utm.my)

This work was supported by the Ministry of Education Malaysia through TRGS under Grant R.J130000.7813.4L844.

ABSTRACT Data collection and storage in a large size is done on a routine basis in any company or organization. To this end, wireless network infrastructure and cloud computing are two widely-used tools. With the use of such services, less time is needed to attain the required output, and also managing the jobs will be simpler for users. General services employ a unique identifier for the aim of storing data in a digital database. However, it may be associated with some limitations and challenges. There is a link between the unique identifier and the data holder, e.g., name, address, Identity card number, etc. Attackers can manipulate a unique identifier for stealing the whole data. To get the data needed, attackers may even eavesdrop or guess. It results in lack of data privacy protection. As a result, it is necessary to take into consideration the data privacy issues in any data digital data storage. With the use of current services, there is a high possibility of exposure and leak of data/information to an unauthorized party during their transfer process. In addition, attacks may take place against services; for instance spoofing attacks, forgery attacks, etc. in the course of information transaction. To address such risks, this paper suggests the use of a biometric authentication method by means of a palm vein during the authentication process. Furthermore, a pseudonym creation technique is adopted to make the database record anonymous, which can make sure the data is properly protected. This way, any unauthorized party cannot gain access to data/information. The proposed system can resolve the information leaked, the user true identity is never revealed to others.

INDEX TERMS Authentication, palm vein, pseudonym, anonimous, unlinkability, data preserving.

I. INTRODUCTION

In current world, all companies and organizations have to hold a huge amount of digital data in storage. The database systems utilized for such purposes need to be properly standardized and well structured in such a way that managing and preserving the data can be done simply. The applied program needs to satisfy the requirements of clients like information holding, personal management, and national statistics, authorization of data tracking, and assistance in administration. In addition, the system used for data preservation must be capable of enhancing the efficiency level of the database record services. In general, data are utilized for the purpose of recording a summarized personal information history that can be shared and retrieved by various users with the help of different online methods.

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Huang.

Though, the systems currently employed by the majority of companies still have problems with data privacy [1]. They make use of a single ID or a unique identifier that represents all attributes of personal records. The records can be linked from a company to another. Finally, many agencies may exchange such records for some certain purposes. For instance, let us assume that an individual is to buy a new number from a telecommunication company by a postpaid method. When his/her status is verified by ID number, she/he is not allowed to complete the purchase since his/her name had been already added to a blacklist by another company because of debt bills. It reveals that any telecommunication company is capable of gaining access to purchasers' status only through the use of a unique identifier. Remember that personal records consist of various information some of which may be of a high sensitivity and confidentiality. If it is observed from the perspective of privacy issue, it can be said that services typically used might result in data exposure. Therefore, the use of such identifiers can be a substantial

security risk. In case information is stolen or lost, any adversary that has access to that information is capable of applying a unique identifier for interfacing all the distinct datasets together to achieve their vicious ends.

The company partnership with various substances is now traceable actively [2]. The privacy challenge has encouraged numerous researchers to search for the most effective ways to make sure of the data record privacy. Currently, literature offers three categories of privacy in general: policy privacy, statistics privacy, and cryptography [3]. It should be noted that biometric cryptography is also appearing as a widely-used method of personal data security. However, the above-mentioned methods have not shown a high capacity in protecting the data privacy [4]. Therefore, there is still a need for efficient biometric methods and processes specifically proposed to protect the data privacy and confidentiality of users. The present study is focused upon creating pseudonyms using a palm vein image on the basis of a pseudonym system in a way to guarantee confidentiality of users and data privacy protection. The palm vein images were taken by means of a palm vein scanner.

Therefore, this paper is extened of [22] which suggests the use of a biometric authentication method by means of a palm vein during the authentication process. Furthermore, a pseudonym creation technique is adopted to make the database record anonymous, which can make sure the data is properly protected. This way, any unauthorized party cannot gain access to data/information. The proposed system can resolve the information leaked, the user true identity is never revealed to others.

The remaining parts of the article is organized as follows. In Section 2, background of the research and related literature are presented and discussed. In Section 3, the biometrics technologies is briefly reviewed. In Section 4, the preserving data privacy approaches are discussed. In Section 5, the proposed method is presented. In Section 6, the discussion is introduced. And finally in Section 7, the whole paper is concluded.

II. BACKGROUND AND RELATED WORKS

The authors in [2] introduced unlinkable pseudonym system for the aim of overcoming the limitations of currently-employed database maintenance systems. As indicated by their findings, data maintenance approaches commonly make use of a unique identifier to make related data sets integrated. The use of commonly-adopted approaches can result in the drawback of controllability. Thus, information is linked simply; any central authority does not exert any form of control or limitation on the information flow. Their system helps to control the information and to share data in a privacy-friendly exchange environment. According to their reports, their proposed system has been verified in the universal composability (UC) framework. They also gave reliable examples on the basis of discreet-logarithm-related assertion. Finally, they have recommended future researchers to comprehensively discuss various techniques existing in literature regarding the way they can offer securely pseudonyms. In [3] a survey was conducted on the privacy-preserving data aggregation specifically within the wireless sensor networks. Existing data protection aggregation schemes were categorized and compared using the most-widely used privacy conservation methods. Findings showed that in the majority of the alternatives, there is a certain phase generally called 'initialization'. During this phase, through a protected channel, participants request main authentication from significant issuers. Thus, there is a need for the development of efficient protocols without depending upon trusted authority and safe two-way communication channels. In another study [5] random cryptographic key was employed for encrypting or encoding the data prior to mining them. The encrypted data play the role of symbols for an unknown dataset. These data are employed to make required decisions with the use of the prepared newly-encoded data mining. Their method was found capable of combining multiparty vertically-partitioned data in a secure way. In addition, a number of parameters such as precision level, utilized memory, error rate, and required time were taken into consideration to assess the efficiency of the technique introduced. Findings showed the optimality of the proposed technique; however, it was still to somewhat time consuming. The researchers in [6] adopted the method of Homomorphic Encryption for the purpose of making sure of the data safety. It offered an acceptable confidentiality level for data since, at any point, the information in this method is not exhibited in plain text. Moreover, the algorithm used in this method shows both efficiency and simplicity. Although optimum results were obtained, the authors recommended the enhancement of the processing data efficiency through lowering the cipher text size. Additionally, a number of algorithms are required to be designed for the aim of searching and querying the encrypted information based on the Full Homomorrphic Encryption (FHE). In [7], an effective scheme, which also made use of the Homomorphic Encryption, was introduced for the protection of data privacy. The designed scheme was found capable of decreasing the overhead communication and the also the amount of energy consumed. Furthermore, this scheme showed a lower probability of data disclosure. Reference [8] proposed a framework to secure data request for cloud and haze computing. The cloud services have been used to test inquired data from haze network when haze network offers inqueried data to clients. In the proposed framework, cloud server predefined some data combination topology trees to haze network, and then haze network may gain associated data from fog nodes according to one of the predefined data combination trees. Furthermore, some haze nodes are allocated as tested nodes which can returns associated data to cloud server. Reference [9] offered a fully ungainly pseudonym modification model to protect privacy in spread networks. Each node utilizes a pseudonym to cary its posts until the ending of the pseudonym, and modifies the pseudonym autonomously after a random delay. Thus, the offered model is established to arrange the spreading of the modify delay, such that at

least k nodes modify their pseudonyms during the delay to pose any enemies. Reference [10] proposed an authentication system to protect senders' privacy in intelligent transportation system. In this system, the vehicles and servers want to connection trusted authority only one time to get secret information and then, based on their secret information, they can produce pseudonyms for the authentication at the receiver side [10].

Also [21] proposed an indirect reciprocity based security framework to guide the behavior of the onboard units in the VANET and reduce the potential attackers and apply the blockchain technique to protect the reputation from being tampered.

III. BIOMETRICS TECHNOLOGY

The current digital era unavoidably demands an extensive application and exchange of digital documents. Almost all companies and organizations, regardless of size and scope, have to make use of certain systems for storing a huge amount of data. It is easier to store and search data compared to conventional techniques in which data were being stored with use of numerous files, then to find required data, users had to search for them in a lot of files. This strategy causes data security to become a key issue. Thus, biometric systems have get popular in terms of data protection from any probable risk [11]. Biometrics systems verify and identify individuals through the use of their biometric traits like behavioral characteristics (e.g., signature and handwriting) or physiological features (e.g., palmprint, fingerprint, palm vein, voice, iris). According to [12], verification process validates that the individual is who she/he claims to be. This process is done through making a comparison between the individual's trait and the biometric profile corresponding to that individual, which exists in database (i.e., one-to-one search). On the other hand, identification actually searches the unknown identities or unknown biometric through making a comparison between the existing biometric characteristic and lots of other features that exist within the database (i.e., one-to-many search).

Every human has their own unique traits. Even twins have the different traits such as their fingerprint. A device is needed in biometric system to capture the traits. Reference [13] maintained that distinctiveness of the human beings' traits helps to capture their specific traits in a precise way by means of sensors and devices capable of transforming data into useful information. Data produced this way are unique and completely applicable to security apps and digital lock that need to have unique access. Biometric traits must be registered in the system prior to the recognition phase. The registration process in this system involves entering the biometric traits, pre-processing, extraction of feature, and finally storing the features extracted. The registration process of biometric traits is depicted in Fig. 1 when the traits are registered in the database, they are then applicable to recognition processes. The recognition process in the biometric system basically involves entering the biometric traits, pre-processing, determining the region of interest, extraction of features,

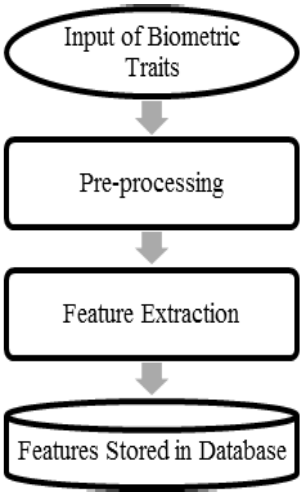

FIGURE 1. Processes involved in registration of biometric features [12].

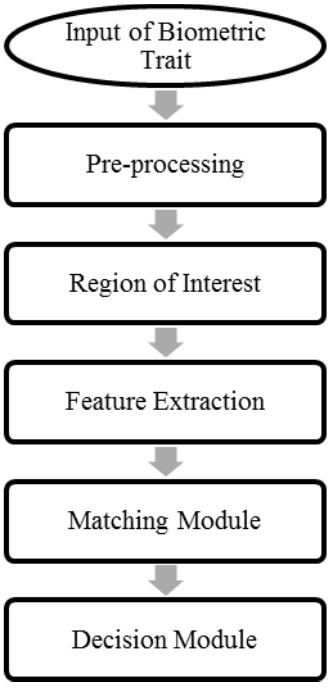

FIGURE 2. Processes involved in the recognition step of the biometric system [12].

matching module, and decision module (see Fig. 2) [12]. In each step of recognition, certain activities need to be performed.

A number of techniques are used in the biometric technology, including palmprint, palm vein, fingerprint, and iris. Each of the above-noted techniques offers some benefits and drawbacks. For instance, palmprint covers an area wider than that of fingerprint, hence containing more characteristics. For that reason, for authentication purposes, palmprint offers a higher reliability compared to fingerprint. In palmprint, there are principal lines, ridges minutia, wrinkles, etc. Additionally, the palmprint pattern of each individual is unique to that person only. Similar to fingerprint, the use of palmprint also

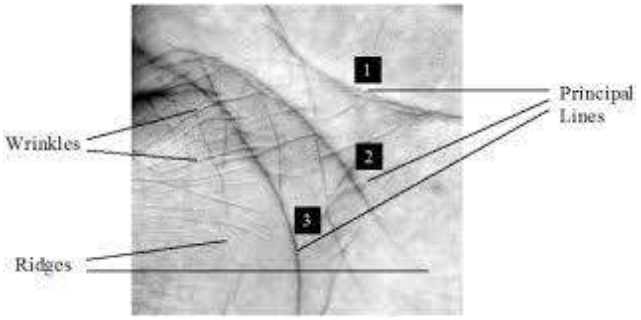

FIGURE 3. Palmprint features [14].

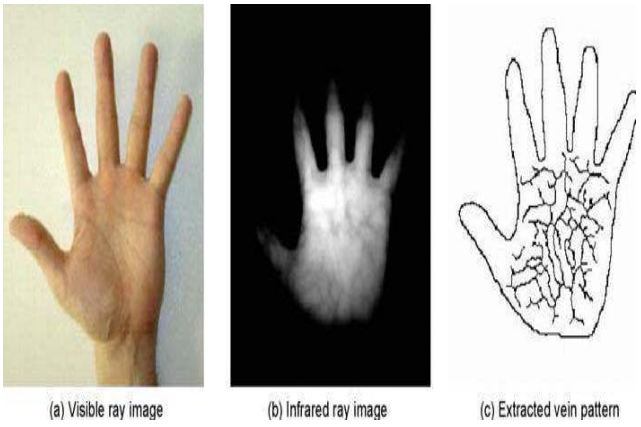

FIGURE 4. Features on the palm vein.

has a number of limitations. It should be also noted that people might leave their palmprint anywhere; thus, it should be also thought as a privacy aspect throughout the authentication processes. On the other hand, irresponsible individuals might leave their palmprint for the purpose of fooling the palmprint scanners. In addition, according to [14]. the database applied in the course of investigation may also affect the effectiveness of palmprint recognition. Those palmprint characteristics that can be properly applied to the authentication and verification systems are depicted in Fig. 3.

As an alternative to palmprint and fingerprint (due to their limitations), palm vein, which has some unique features, has been proposed. Reference [15] maintained that palm vein is not simply altered, damaged, or fabricated. As a result, for identity verification purposes, it can be offered with a high security and reliability. In recent years, systems of palm vein recognition have become very popular since the privacy and security issues have become more significant for companies and organizations. This popularity is also because of the fact that nowadays companies and organizations have to store and manage a huge amount of data. In addition, the palm vein recognition is applicable to authentications and privacy protection purposes, for instance, to gain access to an ATM machine, login into a computer system, gain access to healthcare system, etc. The palm vein popularity is also due to the wider and more complex vascular patterns on individuals' palm, which shows a wealth of distinguishing characteristics. An example of palm vein patterns is presented in Fig. 4.

TABLE 1. Advantages and disadvantages of three biometric traits.

Biometric trait	Advantages	Disadvantages
Fingerprint	Contains reliable features Different pattern for every human being High accuracy for authentication Easy to use Only require small storage space	Quality of images. A person can leave the fingerprints everywhere Can create artificial gummy fingers.
Palmprint	Wider area than fingerprint Consists of more features than fingerprint Different pattern for every person	Quality of images. A person can leave the palmprint everywhere. Can create artificial gummy palm. Database used.
Palm vein	Have unique characteristics Cannot easily damage, change or falsified. Secure and more reliable. Has complicated vascular pattern. Invisible directly by eyes.	The positional of palm vein

Reference [16] believe that intricacy and uniqueness of palm vein patterns have caused these features to offer a highly precise authentication tool. Palm vein has a variety of characteristics that remain persistent all through individuals' lifetime; characteristics that cannot be easily changed or faked. In another study conducted by [17], the biometric technology was comprehensively reviewed and the precision level of each biometric modality was stated roughly. Table 1 summarizes the advantages and disadvantages of fingerprint, palmprint, and palm vein.

According to the above-presented table, palm vein offers more benefits compared to the others. As a result, in the present research, palm vein is applied as the biometric trait to the preservation of data privacy in database management system.

IV. PRESERVING DATA PRIVACY APPROACHES

Literature is consisted of many approaches to preservation of data privacy. On the other hand, data are getting more and more massive and they need to be shared with third parties for specific purposes. In addition, when data are being sharing, they might be disclosed to irresponsible individuals or entities by some destructive attacks. Thus, approaches of a high efficiency and reliability are required for data privacy protection. Reference [18] discussed three currently-adopted approaches to this end. The three approaches are de-identification, privacy preserving aggregation, and operations over encrypted data. Though, the approaches were found infeasible with the lack of anonymity of data. Nevertheless, they concluded that among all, the de-identification approach was the best one regarding the privacy protection in case a highly effective and privacy-protective algorithm can be designed for that purpose. According to [17], lot of approaches have

been introduced by various scholars in order to preserve the privacy of data/information. Though, privacy preserving approaches have not been clearly classified yet. For that reason, [17] classified the data-preservation approaches into two categories: cryptographic and non-cryptographic. On the other hand, adversaries also make use of new technology and always show their own progressive techniques for collecting data/information in illegal ways. As a result, [17] confirmed the pressing need for designing robust algorithms in this regard.

A. PSEUDONYM SYSTEM

With the use of pseudonym systems, which was proposed by [19], users will be allowed to have interaction with numerous organizations anonymously. This system effective in an effective and anonymous way. Each organization might know one user through different pseudonyms. These nyms cannot be linked, which guarantees the data anonymity. In this system, processes are applied to the interactions between user and corresponding organization. The processes involved in this system are:

1. Producing master key: In this initial phase, master key pairs are produced for the user and the organization. Here, 'user' can refer to an individual, a group of people, or company/organization. Each user possesses a master secret key corresponding to a master public key. The key pairs are capable of corresponding to the digital signature created by a user to sign documents or receive encrypted data.
2. Registering with the CA organization: This organization keeps true identities of users and knows the master public key of each user. CA is responsible for guaranteeing that all users possess a master secret key and a master public key, which will be compromised in case the user cheats. The user's master public key is actually the user's nym in interactions between the user and CA. CA is able to confirm the validity of the user.
3. Registering with an organization: For a nym to be generated for a certain user, the user needs to contact the organization; then, they can together compute the nym. The master public/secret key pairs of users are extracted by an identity extractor. After that, the user can demonstrate to the organization that s/he possesses a credential generated by CA in order to authenticate him or herself.
4. Issuing credentials: For achieving a credential, an interactive protocol is employed to engage the user and the organization.
5. Transferring the credentials: when a user receives valid credentials, s/he will be able to make interaction with organizations without revealing true identity. Such process is termed 'transfer of a credential' since in this procedure, a credential is transferred from the user's pseudonym with an organization towards his/her pseudonym with another organization. The key objective of the pseudonym system is countering two types of widely-observed attacks, i.e., credential forgery and user identity compromise or pseudonym linking.

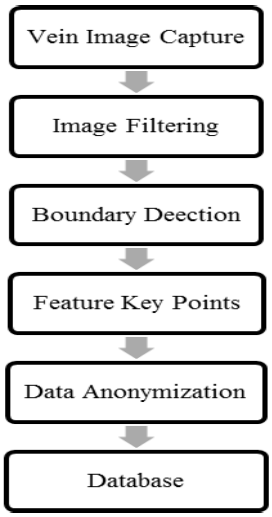

FIGURE 5. Propsed data preserving methodology.

V. PROPOSED METHOD

The present paper is mainly aimed to utilize the pseudonym technique for the purpose of preserving data privacy through the use of palm vein. palm vein, as mentioned earlier, is type of reliable biometric characteristic that can be applied to the generation of pseudonyms. For the generation of pseudonym, the method of homomorphic encryption is adopted. Fig. 5 summarizes the proposed privacy preserving methodology. To take the palm vein image, the palm vein scanner is utilized in this study. The images captured are typically in need of filtration in a way to remove any noise. The, the edge detection method is used to explore the feature boundary. A number of points in the selected feature are taken into account as the vein signals that can be properly translated into the secret key. Two partial keys are generated from the palm vein in order to assure the proposed technique is capable of protecting the data privacy against probable adversary attacks. Next, the partial keys produce the final pseudonym.

1) VEIN IMAGE CAPTURE

Vein images are taken by palm vein scanner or gathered from palm vein database. An image taken by the palm vein scanner is depicted in Fig. 6.

2) IMAGE FILTERING

To have a higher quality image, the captured image should be processed with filtering tools. It is done using the Median Filter method in order clear out the noise along with Contrast Limited Adaptive Histogram Equalization (CLAHE). Generally, the nonlinear method of Median filtering is applied to the removal of ''salt and pepper'' noise of images. In cases the objective of filtering is decreasing noise and, at the same

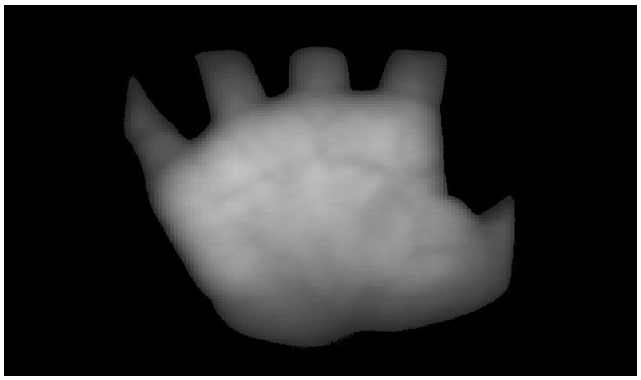

FIGURE 6. Pam vein image.

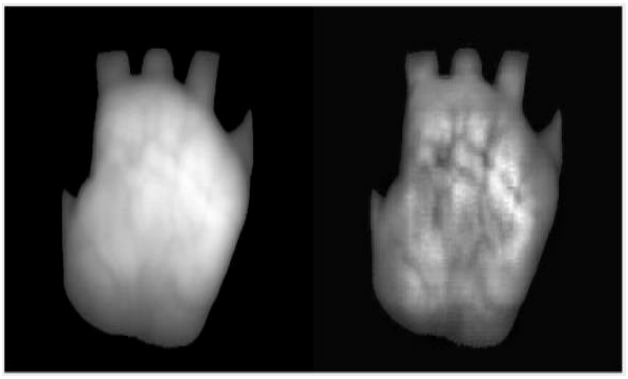

FIGURE 7. Left (original image) right (filter image).

time, maintaining edges, a median filter offers a higher level of efficiency compared to convolution.

CLAHE works on small areas of the image, or tiles, rather than the entire image as one entity. In this method, each tile's contrast is increased in a way to roughly match the histogram of the output region that is specified by a certain value. After that, the adjacent tiles are integrated aiming at removing artificially generated boundaries with the use of bilinear interpolation. Contrast can be confined to evade the amplification of any noise that might exist within the image, particularly in homogeneous regions. The result of an image filtering process is displayed in Fig.7.

3) BOUNDARY DETECTION

Subsequent to filtering, the image boundary is detected by means of Canny edge detector. This technique is adopted for the purpose of exploring the vein pattern. The result obtained by a Canny edge detector is presented in Fig. 8.

According to [20], the Canny edge detection operator, which is a multi-stage edge detection algorithm, was designed by John in 1986. This algorithm takes into consideration three criteria to evacuate the merits of the detecting operator: a high SNR criterion, a single-edge single response criterion, and a precise positioning criterion. The algorithm is composed of three phases: de-noising, searching for a gradient of light, and following the edge. With taking into consideration the complicated computation of the standard algorithm, the processing protocol is simplified as follows:

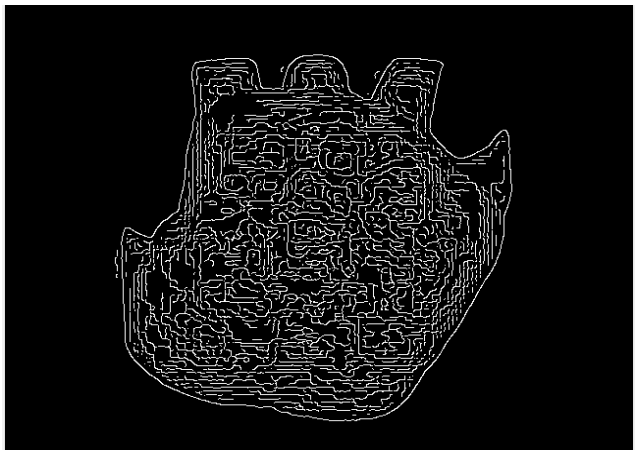

FIGURE 8. Result of using canny detector.

Phase 1: As original images contain noise, a direct edge detection might negatively affect the edge detection algorithm operations. As a result, the Gaussian smoothing filtration of the original image needs to be performed in such a way that the phase noise of a single pixel cannot affect the rim detection. Practically, eight pixels close to the filtration point are averaged to achieve a smooth image.

Phase 2: Sobel operators are applied in various directions with points to be measured (point 8 in the vicinity) for the purpose of determining the gradients in 45, 135, horizontal and vertical directions. For this point, the maximum of the four gradients is taken as the gradient, while the maximum gradient direction will be reported, hence achieving a brightness gradient map, which includes directional data.

Phase 3: High and low thresholds are used to identify edges. The point with a higher gradient value is probably belonged to the edge of the brightness gradient map; however, no simple way exists for choosing the threshold. At first, the higher threshold is applied to the detection of the brightness gradient attained in order to gain a reliable boundary contour with a low consistency. After that, the low threshold is applied to tracking the search alongside the edge of the contour, which is determined by the high threshold. It results in the identification of the boundary of the fuzzy region.

4) FEATURE KEY POINTS

The Harris Stephens method, which was proposed for detecting the corner point, was applied to the extracted image feature. Controlling the picture was done using the windows average function part of this process. Remember that the pattern of the blood stream differs for every individual. With use of the above-mentioned process, the blood vessel pattern can be explored and traced in an efficient way. The vein key points are displayed in Fig.9.

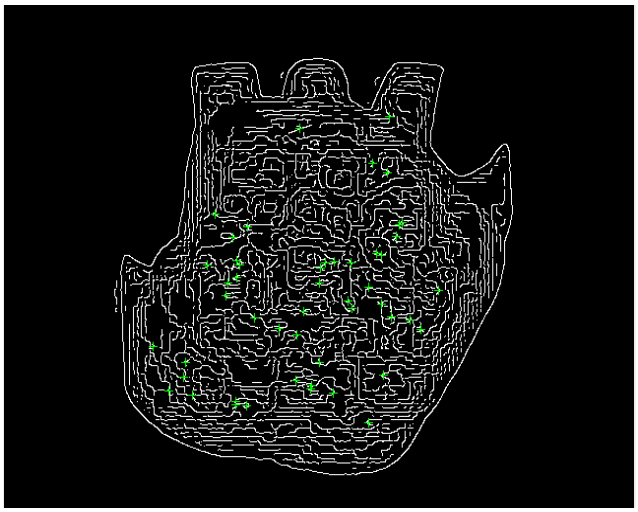

FIGURE 9. Feature key points.

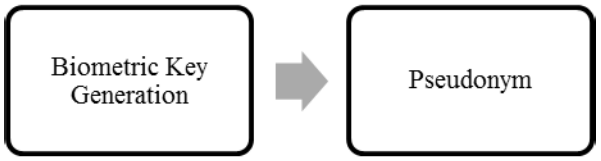

FIGURE 10. Pseudonym generation.

5) DATA ANONYMIZATION

Anonymization is an attempt made for preserving the publishing data privacy and protecting sensitive information and owners identity. In this regard, the pseudonym system has been also introduced. With the use of a digital pseudonym, a public key can authenticate an anonymous sender's signature, who signed a message using his/her own private key. Though, it does not have the adequate capacity that is necessary to be well protected against eavesdrop attackers. Furthermore, the unlinkable functionality is not adequate for the proper protection of privacy of the consumers' information. For that reason, the pseudonym system still needs to be further improved. The flow to produce pseudonym for data anonymization is depicted in Fig. 10.

6) BIOMETRIC KEY GENERATION AS MASTER KEY

At this stage, following the collection of the feature key points, a biometric key is produced as the master key. When the function is extracted and the edges are chosen as key feature points, then the palm vein image matrix is used aiming for the measurement of the Euclidean distance metric for the coordination of the distance of an image component with a mathematical expression in m and n plane, as follows:

$$d^2_{m,n} = (m_1 - n_1)^2 + (m_2 - n_2)^2$$

Next, with neighborhood operation, more processing is done on the measured Euclidean distance matrix in a way to choose some values of the Euclidean distance matrix computed. After that, the chosen distance matrix stream will be converted to ' 0 and 1 ' bit stream. For the purpose of converting to a decimal number, the bit of the chosen stream should be processed. The decimal number produced is the key to the specific palm vein image sample, hence being also the key to the biometric trait. To produce the key, the following steps need to be taken into action:

TABLE 2. Selected thirty elements in a row.

0.97061348	**0.30819321**	**0.6152229**	**0.97224426**
0.97061348	0.96878052	0.96670341	0.96433163
0.9615955	0.95840549	0.95463943	0.95012474
0.94461441	0.93774223	0.92893219	0.91723728
0.90098047	0.87689447	0.83772230	0.76393199
0.58578646	0	0	0
0	0.58578646	0.76393199	0.17157292
0.83772230	0.39444876		

TABLE 3. Thirty bit stream binary vector.

1	0	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
0	0	0	0	1	1	0
1	0					

1. From a computed Euclidean 128 × 128 matrix of palm vein image processed by means of image filter and edge detection system, the middle thirty elements within a row are chosen. An example of the elements chosen this way is presented in Table 2.
2. As can be observed in Table 3, the thirty-bit stream binary vector is formed through applying the round off function to the values from Table2, which are chosen as row values of the computed Euclidean distance matrix of the palm vein images.
3. After that, an algorithm is used for the aim of converting the binary stream of thirty bits into a decimal number as "805305882."

1011 1111 1111 1111 1111 1000 0110 10 = 805305882

4. It forms a decimal number from a binary vector of thirty bit, and this will act as a powerful biometric key or master key that can be properly applied to recognizing the subject.

7) PSEUDONYM

To obtain the pseudonym, the ROI of palm vein must create two partial pseudonyms. First pseudonym is generated from the fetures keys and secon partial pseudonym generate from random numbers of user ID. After this, both of the

partial pseudonym will be merged into the final pseudonym made. At this stage, HMAC-SHA256 is used to collect final pseudonym.

The authors in [12] pioneered the idea of pseudonyms production. In this method, a one-time pseudonym is created through the application of a keyed-hash message authentication code (HMAC) to secure information of a user who has already registered on the system. HMAC-SHA256 method is used for hashing. HMACSHA256 is a type of keyed hash algorithm that is built from the SHA-256 hash function and used as a HMAC. The HMAC method mixes the secret key with the message data, hashes the result with the hash function, re-mixes the hash value with the secret key, and then applies the hash function a second time. The output hash is 256 bit long. The Cloud Server (CS) refers to a computer used for producing the pseudonyms. A pseudonym is created during a six-step process as explained in the following:

1. The process starts with a Random Number (R) which is generated by Trusted Authority Identification (TA ID).
2. CS generates HMAC by means of (TA ID) and master key (K) through a concatenation procedure. In case this master key is uncovered, then the real name security will be compromised.
3. The cryptographic HMAC is applied to the ID for N times in order to create a one-time pseudonym hash chain. The application of the cryptographic hash function will lead to formation of Pseudonyms (PIDs).
4. At this step, the initial secret ID will be discarded.
5. Then, the database is supplied with N pseudonyms.
6. PID1 . . . PIDend will be discarded from CS. Only the HMAC (ID|| K), the one at the top of the user list, will be stored on the file.

VI. PRIVACY PRESERVING ANALYSIS

The PID published on the database is special and unlike, apart from the fact that the user only knows his PID when it was shown on the information site, the identity of another user can not be identified in a way. This requirement is achieved because the pseudonym shown on databases is in series order. Users only learn when they look at TA's list of pseudonyms submitted. Thus the user can not be aware of another user's identity. Table 4 shows the comparison of the previous works and proposed method.

VII. DISCUSSION

The method proposed here is robust enough to provide an appropriate protection of data privacy against system attacks, e.g., forgery attacks, robbery attacks, insider attacks, spoofing attacks, and offline password guessing. The method has the capacity of addressing the issues in relation to data privacy preservation, e.g., information leakage. The biometric trait of palm veins has been found of a higher accuracy level compared to other traits, which is due to the fact that palm vein has unique features. The palm vein pattern is hardly altered or faked. As a result, this trait offers safety and reliability in the process of confirming individuals' identity. The proposed pseudonyms production can enhance the privacy of data and protect them against any hazard. The use of unlinkable feature ensures that unauthorized parties would not be capable of predicting the data owner. This pseudonym differs for every information a user has.

TABLE 4. Comparison of previous works and propose method.

Year	Method	Drawback
2015 [2]	(Un)linkable Pseudonym System	Drawback of controllability.
2015 [3]	Data Protection Aggregation Schemes	Depending upon trusted authority and less safe for two-way communication channels
2016 [5]	Random cryptographic key	Time consuming
2016 [6]	Homomorphic Encryption	Processing data efficiency
2020	Proposed Method	The user can not be aware of another user's identity. Information of user is anonymous.

VIII. CONCLUSION

Today, all companies and organizations have to store a massive quantity of data and protect them properly. To have such data effectively handled, the cloud computing services are employed widely. For the data protection purposes, the present study is focused upon anonymized data. An innovative method was proposed for generating pseudonym in a way to make sure of preserving the information security. In addition, the solutions proposed can improve the security of data in different systems.

REFERENCES

[1] N. M. Shrestha, A. Alsadoon, P. W. C. Prasad, L. Hourany, and A. Elchouemi, "Enhanced E-health framework for security and privacy in healthcare system," in *Proc. 6th Int. Conf. Digit. Inf. Process. Commun. (ICDIPC)*, Apr. 2016, pp. 75–79.

[2] J. Camenisch and A. Lehmann, "(Un)linkable pseudonyms for governmental databases," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, 2015, pp. 1467–1479.

[3] J. Xu, G. Yang, Z. Chen, and Q. Wang, "A survey on the privacy-preserving data aggregation in wireless sensor networks," *China Commun.*, vol. 12, no. 5, pp. 162–180, May 2015.

[4] J.-J. Yang, J.-Q. Li, and Y. Niu, "A hybrid solution for privacy preserving medical data sharing in the cloud environment," *Future Gener. Comput. Syst.*, vols. 43–44, pp. 74–86, Feb. 2015.

[5] S. Sharma and D. Shukla, "Efficient multi-party privacy preserving data mining for vertically partitioned data," in *Proc. Int. Conf. Inventive Comput. Technol. (ICICT)*, Aug. 2016, pp. 1–7.

[6] M. M. Potey, C. A. Dhote, and D. H. Sharma, "Homomorphic encryption for security of cloud data," *Procedia Comput. Sci.*, vol. 79, pp. 175–181, Mar. 2016.

[7] V. Akila and T. Sheela, "Preserving data and key privacy in data aggregation for wireless sensor networks," in *Proc. 2nd Int. Conf. Comput. Commun. Technol. (ICCCT)*, Feb. 2017, pp. 282–287.

[8] K. Gu, N. Wu, B. Yin, and W. J. Jia, "Secure data query framework for cloud and fog computing," *IEEE Trans. Netw. Service Manage.*, to be published.

[9] Z. Liu, L. Zhang, W. Ni, and I. Collings, "Uncoordinated pseudonym changes for privacy preserving in distributed networks," *IEEE Trans. Mobile Comput.*, to be published.
[10] C. N. H. Vinh, A. Truong, and T. T. Huu, "A privacy preserving authentication scheme in the intelligent transportation systems," in *Proc. Int. Conf. Future Data Secur. Eng.* Cham, Switzerland: Springer, 2018, pp. 103–123.
[11] P. Balakumar and R. Venkatesan, "A survey on biometrics based cryptographic key generation schemes," *Int. J. Comput. Sci. Inf. Technol. Secur.*, vol. 2, no. 1, pp. 80–85, 2012.
[12] M. O. Oloyede and G. P. Hancke, "Unimodal and multimodal biometric sensing systems: A review," *IEEE Access*, vol. 4, pp. 7532–7555, 2016.
[13] I. Verma and S. K. Jain, "Biometrics security system: A review of multimodal biometrics based techniques for generating crypto-key," in *Proc. 2nd Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, 2015, pp. 1189–1192.
[14] A. Harb, M. Abbas, A. Cherry, H. Jaber, and M. Ayache, "Palm print recognition," in *Proc. Int. Conf. Adv. Biomed. Eng. (ICABME)*, Sep. 2015, pp. 13–16.
[15] K.-S. Wu, J.-C. Lee, T.-M. Lo, K.-C. Chang, and C.-P. Chang, "A secure palm vein recognition system," *J. Syst. Softw.*, vol. 86, no. 11, pp. 2870–2876, Nov. 2013.
[16] S. D. Raut and V. T. Humbe, "Review of biometrics: Palm vein recognition system," *IBMRDs J. Manage. Res.*, vol. 3, no. 1, pp. 217–223, 2014.
[17] J. A. Unar, W. C. Seng, and A. Abbasi, "A review of biometric technology along with trends and prospects," *Pattern Recognit.*, vol. 47, no. 8, pp. 2673–2688, Aug. 2014.
[18] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," *IEEE Netw.*, vol. 28, no. 4, pp. 46–50, Jul. 2014.
[19] D. Chaum, "Security without identification: Transaction systems to make big brother obsolete," *Commun. ACM*, vol. 28, no. 10, pp. 1030–1044, Oct. 1985.
[20] Q. Pu, J. Wang, and R. Zhao, "Strong authentication scheme for telecare medicine information systems," *J. Med. Syst.*, vol. 36, no. 4, pp. 2609–2619, 2012.
[21] C. Dai, X. Xiao, Y. Ding, L. Xiao, Y. Tang, and S. Zhou, "Learning based security for VANET with blockchain," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, Dec. 2018, pp. 210–215.
[22] N. H. M. Nazari and S. A. Razak, "Palm vein pseudonym for anonymous database record," in *Proc. IEEE Conf. Appl., Inf. Netw. Secur. (AINS)*, Nov. 2018, pp. 38–42.

SHUKOR ABD RAZAK is currently an Associate Professor with Universiti Teknologi Malaysia. He also actively conducts several types of research in digital forensic investigation, wireless sensor networks, and cloud computing. He is the author or coauthor of many journals and conference proceedings at national and international levels. His research interests are on the security issues for mobile ad hoc networks, mobile IPv6, vehicular ad hoc networks, and network security.

NUR HAFIZAH MOHD NAZARI was born in Taiping, Perak, Malaysia. She received the B.S. degree in computer science (bioinformatics) from Universiti Teknologi Malaysia, in 2016, where she is currently pursuing the M.Phil. degree in computer science with the Information Assurance and Security Research Group (IASRG), under the supervision of Assoc. Prof. Dr. S. Razak.

ARAFAT AL-DHAQM was born in Mukhayliya, Lahij, Yemen, in 1976. He received the B.S. degree in information systems from the University Technology of Iraq, in 2002, and the master's and Ph.D. degrees in computer science (digital forensics and information security) from University Technology Malaysia, in 2013 and 2018, respectively, where he is currently pursuing the Ph.D. degree under the supervision of Assoc. Prof. Dr. S. Razak.